DTIC® DEFENSE TECHNICAL INFORMATION CENTER

Information for the Defense Community

DTIC® has determined on 03 / 20 / 2015 that this Technical Document has the Distribution Statement checked below. The current distribution for this document can be found in the DTIC® Technical Report Database.

☑ **DISTRIBUTION STATEMENT A.** Approved for public release; distribution is unlimited.

☐ **© COPYRIGHTED.** U.S. Government or Federal Rights License. All other rights and uses except those permitted by copyright law are reserved by the copyright owner.

☐ **DISTRIBUTION STATEMENT B.** Distribution authorized to U.S. Government agencies only (fill in reason) (date of determination). Other requests for this document shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT C.** Distribution authorized to U.S. Government Agencies and their contractors (fill in reason) (date determination). Other requests for this document shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT D.** Distribution authorized to the Department of Defense and U.S. DoD contractors only (fill in reason) (date of determination). Other requests shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT E.** Distribution authorized to DoD Components only (fill in reason) (date of determination). Other requests shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT F.** Further dissemination only as directed by (insert controlling DoD office) (date of determination) or higher DoD authority.

Distribution Statement F is also used when a document does not contain a distribution statement and no distribution statement can be determined.

☐ **DISTRIBUTION STATEMENT X.** Distribution authorized to U.S. Government Agencies and private individuals or enterprises eligible to obtain export-controlled technical data in accordance with DoDD 5230.25; (date of determination). DoD Controlling Office is (insert controlling DoD office).

THE UNIVERSITY OF MICHIGAN

COLLEGE OF ENGINEERING
ATMOSPHERIC, OCEANIC & SPACE SCIENCES

SPACE RESEARCH BUILDING
2455 HAYWARD STREET
ANN ARBOR, MICHIGAN 48109-2143

October 14, 2014

Melinda Peng
Code 7530
7 Grace Harper Avenue, Stop 2
Monterey, CA 93943-5502

Subject: Final Progress Report Grant #N00173-10-1-G035

On behalf of the Principal Investigator, Derek Posselt, and in compliance with the requirements of the Grant No. N00173-10-1-G035 titled, "Toward an Operational Particle Filter-Based Ensemble Data Assimilation System", I am forwarding the enclosed hard copy of the final progress report.

If you have any questions or need additional information please contact Dr. Posselt at (734)936-0502.

Sincerely,

Cheri Johnson

Cheri Johnson
Research Administrator

Encls.

cc: Defense Technical Information Center.
Naval Research Lab, Code 5596
Chicago Regional Office, Code N62880
File F026067

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*: 09/22/2014

2. REPORT TYPE: Final Technical Report

3. DATES COVERED *(From - To)*: 9/1/2010-5/15/2014

4. TITLE AND SUBTITLE
Toward an Operational Particle Filter-Based Ensemble Data Assimilation System

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00173-10-1-G035

5c. PROGRAM ELEMENT NUMBER
NRL BAA 75-09-01

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Posselt, Derek

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Michigan
3003 S. State St.
Ann Arbor, MI 48109-1274

8. PERFORMING ORGANIZATION REPORT NUMBER
F026067

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Chicago Regional Office
230 South Dearborn, Room 380
Chicago, IL 60605-1595

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The primary goal of this project was to use a Markov chain Monte Carlo (MCMC) algorithm to examine the strengths and limitations of ensemble data assimilation algorithms, when applied to estimation of convective cloud system properties. MCMC methods return an exact solution, and as such are powerful tools for the evaluation of approximate data assimilation algorithms. The research was successful in advancing ensemble data assimilation theory, and has directly led to two peer reviewed journal articles, a peer reviewed book chapter, and 11 conference presentations.

15. SUBJECT TERMS
Data Assimilation, Ensemble Kalman Filter, Markov chain Monte Carlo, Clouds and Precipitation, Convection

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Derek Posselt
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	6	734-936-0502

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATE COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33315-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report. e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Toward an Operational Particle Filter-Based Ensemble Data Assimilation System

Final Technical Report

Grant Number N00173-10-1-G035
NRL BAA 75-09-01 Atmospheric Effects, Analysis, and Prediction

PI: Dr. Derek Posselt

Department of Atmospheric, Oceanic, and Space Sciences
2455 Hayward Street
University of Michigan, Ann Arbor, MI
Phone: (734) 936-0502
Email: dposselt@umich.edu

1. Research Background and Motivation

Data assimilation addresses the dual goals of (1) producing the best estimate of the state of the physical system (e.g., atmosphere, ocean, land) and (2) quantifying the uncertainty on that estimate. All modern data assimilation systems are based on the idea that the most complete information about the state of a system can be expressed as a conjunction of information contained in a prior estimate, a set of observations, and a model of the physical system. Probability distribution functions (PDFs) are used to define each type of information, and the solution is obtained from the joint probability distribution of the control variable(s) and the available information. Over the past 50 years, research in statistical estimation techniques and their application to large systems has resulted in estimates of the atmosphere, ocean, and land state that are robust for most cases. However, significant challenges remain to be addressed. Key among these are how to properly account for model error and how to produce estimates for highly nonlinear systems, especially for high-impact weather events (e.g., severe storms).

Ensemble data assimilation algorithms have increasingly been used in nonlinear systems because they do not require use of a linear approximation to the forecast model (e.g., tangent linear model or adjoint). Ensemble assimilation algorithms produce a solution by generating a sample of the joint PDF of interest, but are subject to potentially limiting assumptions about these probabilities (e.g., Gaussian). By contrast, Markov chain Monte Carlo (MCMC) algorithms require no specific form of the probability distributions of interest, and produce a sample of the true solution probability. MCMC has been used to effectively characterize uncertainty and information content in remote sensing retrievals (Posselt et al., 2008; Posselt and Mace, 2014), and to assess the uncertainty in model physics parameterizations (Posselt and Vukicevic, 2010). MCMC algorithms are well-suited to non-Gaussian estimation in the presence of nonlinearities, but are relatively computationally expensive and are only practical for relatively simple models and low-dimensional systems. The research conducted in the course of this project was designed to use MCMC to evaluate two modern data assimilation systems developed at the Naval Research Laboratory in Monterey, CA. In the process, it seeks to move the data assimilation community in the direction of an operational particle filter-based data assimilation system.

The following tasks were conducted in the course of the three-year project:

1. Examine the strengths and limitations of Ensemble Kalman Filter (EnKF) type data assimilation algorithms when applied to estimation of convective cloud system properties.
2. Determine whether EnKF algorithms are capable of representing rapid changes in the state associated with transitions between convective and stratiform precipitation.
3. Explore the degree to which EnKF techniques can properly return estimates of positive definite quantities (e.g., cloud content).
4. Assess whether a recent innovation on traditional EnKF algorithms (the quadratic filter; Hodyss, 2011) is capable of improving the representation of positive definite quantities.

Each of these three tasks draws on the PI's expertise with nonlinear data assimilation methods, and leverages the resources and expertise available at the Naval Research Laboratory in Monterey, CA. In the following sections, the results obtained in each of the above task areas is briefly described, along with reference to relevant publications.

2. Estimation of Convective Cloud System Properties Using an EnKF Algorithm

This work was designed to assess the effectiveness of an Ensemble Transform Kalman Filter (ETKF) to represent convective processes. Previous research found that the probability density functions (PDFs) of cloud microphysical parameters were non-Gaussian, and in many cases, had a non-unique solution (multiple PDF modes). In this work, we built upon the work of Posselt and Vukicevic (2010), and used MCMC to examine the degree to which an ETKF algorithm was able to characterize non-Gaussian PDFs. We utilized a column convective model and built an ETKF algorithm suitable for performing model parameter estimation.

The major findings of this research consisted of the following:

1. The effect of model nonlinearity on the posterior PDF is included in ETKS in that the nonlinear model is allowed to respond to changes in model parameters; the full nonlinear model propagates perturbations forward in time. However, these changes are constrained by the requirement that the ETKS posterior perturbations be strictly linear functions of the prior perturbations. In contrast, the accept-reject procedure of the MCMC finds posterior perturbations that can be any function of the prior perturbations.

2. Ensemble Kalman Smoothers can preserve key aspects of Non-Gaussian priors. Specifically, the ETKF was found to give a qualitatively accurate multi-modal posterior PDF when given an accurate multi-modal prior. The implication is that it is not the Gaussian assumption used in the derivation of the ETKS that causes mis-representation of the posterior. Instead, it is the lack of information on higher moments and/or multiple modes in the prior ensemble.

3. Response of ETKF Estimates to Changes in Convective Regime

In this portion of the research, we used the same column model framework described in section 2 to explore the extent to which EnKF type algorithms are capable of tracking changes in cloud regime with time. Specifically, the model is designed to simulate a transition between deep convection and stratiform rainfall half way through its three-hour integration. The PDFs of cloud microphysical variables change significantly at this transition point due to the influence of different parameters at different stages of the convective life cycle. MCMC naturally tracks the effect of these changes on the model output, but it was unclear whether an EnKF algorithm is capable of doing the same. We generated posterior probability distributions using sequentially greater numbers of observations in time, and evaluated the efficacy of the EnKF as compared with MCMC.

The major conclusions of this study were the following:

1. Ensemble Kalman smoothers perform poorly when the posterior mean and perturbations are strongly non-linear functions of the forecast error. This was evidenced by a failure of the ETKF to represent the transition of the posterior PDF from a uni-modal to multi-modal form when observations of the stratiform phase of cloud evolution were assimilated. Though the ETKS is unable to produce a multimode analysis from a uni-mode prior, the posterior PDF produced by both the deterministic and perturbed observations versions of the ETKS is clearly non-Gaussian.

2. The uncertainty characteristics of model physics parameterizations depend critically on the characteristics of the environment. Abrupt changes in the physical environment can lead to similarly abrupt changes in parameter uncertainty. Ensemble Kalman Filter-type algoritms, while not capable of capturing rapid (nonlinear) transitions in the nature of the posterior PDF, can be shown to perform well if provided with a robust prior ensemble. As such, Ensemble-Kalman-Filter-type algorithms have promise as cost-efficient methods for model parameter estimation.

4. Determination of whether ETKF Algorithms can Represent Positive Definite Quantities

This research addressed the question of whether ensemble filters, which employ the full nonlinear model, are capable of representing quantities that are hard bounded. In this case, the quantities of interest are cloud microphysical parameters that are hard bounded at a value of zero. As in the research described in sections (2) and (3) above, this work employed a column convection model with control parameters that were tunable constants in the cloud microphysical scheme. In contrast to previous work, which examined only one set of microphysical values, this work tested several sets of values. The goals were to determine (1) how the posterior PDF changes with true state, and (2) whether the EnKF estimate degrades with proximity to a hard bound.

The major conclusions of this work were the following:

1. The true analysis ensemble, as constructed from samples of the Bayesian posterior distribution, changes shape significantly with changes in the true parameter set for a model in which control parameters are nonlinearly related to the observations.

2. Multimodality is realized only in certain regions of the parameter space, and is associated with non-monotonicity in the parameter-observation response function.

3. EnKF algorithms produce PDFs with increasing probability mass at non-physical values as parameter values approach zero. In fact, for parameters very close to zero, the posterior mean may be non-physical.

4. The slope of the parameter-observation response function determines parameter sensitivity, and, by extension, the posterior variance. A constant response function derivative leads to posterior variance that is independent of the true parameter value. This is consistent with results found by Hodyss (2011), who showed that the first derivative of the response function with respect to observations determines the posterior variance while the second derivative determines the posterior third moment.

5. Evaluation of a Quadratic Ensemble Filter

In this final portion of the work, we explored the extent to which an ensemble filter that accounts for skewness in the PDFs of interest can improve the solution PDF for nonlinear and non-Gaussian quantities. We implemented a version of the Quadratic Ensemble Filter (QEF),

developed at NRL-Monterey, and tested it for the same convective system used in topics (2) – (4) above. The major conclusions of this work were the following:

1. The true error distribution for a given set of observations (the "error of the day") is not reproduced by ensemble smoothers. Instead, the distribution produced by ensemble filters (ETKF and QEF) consists of the expected analysis error covariance matrix produced by integrating over all possible observations for a given prior.

2. When approximate ensemble solutions are compared with the integral over multiple Bayesian posteriors constructed from multiple draws of the true parameters from the prior (e.g., by running MCMC multiple times using different true parameter sets), both ETKF and QEF can be shown to provide a realistic estimate of the average posterior analysis distribution, but with larger ensemble variance than that of the average of the true posterior analysis distributions.

3. A filter constructed with a nonlinear update that accounts for the effects of skewness in the prior and posterior distribution produces on average an estimate that is more consistent with the true posterior ensemble mean, but which still fails for cases with non-monotonic nonlinearity. For these cases, the mean is closer to the true mean than an EnKS algorithm, but, like the EnKS, is also not restricted to regions of phase space where known physically consistent solutions exist. In addition, for state estimates hard bounded at some value (e.g., zero for concentrations of scalar quantities), a significant portion of the posterior ensemble density may lie in an unphysical region of the parameter space. This becomes more marked when the observed concentrations and/or parameter values approach the specified limit.

6. Summary

This project led to significant advances in the understanding of ensemble data assimilation theory, and has subsequently supported the development of new assimilation schemes more suitable for nonlinear systems. MCMC was shown to be a robust tool for the evaluation of ensemble data assimilation schemes, and exploration of how MCMC may be used to evaluate new state of the art assimilation systems is now underway.

Peer Reviewed Publications Produced in Support of this Project

Posselt, D. J., and C. H. Bishop, 2012: Nonlinear parameter estimation: Comparison of an Ensemble Kalman Smoother with a Markov chain Monte Carlo algorithm. *Mon. Wea. Rev.*, **140**, 1957-1974.

Posselt, D. J., 2013: Markov chain Monte Carlo Methods: Theory and Applications. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications, 2nd Ed.* S. K. Park and L. Xu, Eds., Springer, pp 59–87.

Posselt, D. J., D. Hodyss, and C. H. Bishop, 2014: Errors in Ensemble Kalman Smoother Estimates of Cloud Microphysical Parameters, *Mon. Wea. Rev.*, **142**, 1631-1654.

Conference Presentations Directly Related to this Project

Posselt, D. J., *Use of Markov Chain Monte Carlo Algorithms for Model Uncertainty Estimation: Beating the Problem of Nonuniqueness.* Invited talk presented at the American Geophysical Union Fall Meeting, San Francisco, CA. December 2010

Posselt, D. J., *On the Use of Data Assimilation Methods to Quantify Uncertainty in Model Physics Parameterizations,* Poster presented at the World Climate Research Program Open Science Conference, Denver, CO. October 2011.

Posselt, D. J., *On the use of data assimilation methods to quantify uncertainty in model physics parameterizations*, Poster presented at the American Geophysical Union Fall Meeting, San Francisco, CA. December 2011.

Posselt, D. J., *Nonlinear Model Parameter Estimation: Comparison of Results From a Markov Chain Monte Carlo Algorithm and An Ensemble Transform Kalman Filter*, Talk presented in the 16th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS) at the 92nd American Meteorological Society Annual Meeting, New Orleans, LA. January 2012.

Posselt, D. J., *Quantifying Uncertainty in Model Physical Parameterizations,* Poster presented at the 1st Pan-GASS Workshop, Boulder, CO. September 2012.

D. J. Posselt, *A Markov Chain Monte Carlo-Based Examination of the Interaction Between Model Physics Uncertainty and Model State*. Poster presented at the American Geophysical Union Fall Meeting, San Francisco, CA. December 2012.

D. J. Posselt, *Simultaneous Nonlinear Estimation of Model Physics Uncertainty and Model State in Simulations of Deep Convection*, Talk presented in the 17th Conference on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS) at the 93rd American Meteorological Society Annual Meeting, Austin, TX. January 2013.

Posselt, D. J., *Model Error Analysis: Uncertainty Inherent in Model Physics Parameterizations*, Invited talk presented at the Society for Industrial and Applied Mathematics Annual Meeting, San Diego, CA. July 2013.

Posselt, D. J., *On the Use of Data Assimilation Methodologies for Examining Cloud System - Environment Interactions*, Talk presented at the 6th World Meteorological Organization Symposium on Data Assimilation, College Park, MD, October 2013.

Posselt, D. J., D. Hodyss, and C. H. Bishop, *Assessing Ensemble Filter Estimates of the Analysis Error Distribution of the Day*, Poster presented at the 2013 American Geophysical Union Fall Meeting, San Francisco, CA, December 2013.

Posselt, D. J., *On the Use of Data Assimilation Methodologies for Examining Cloud System - Environment Interactions*, Invited talk presented at the 2013 American Geophysical Union Fall Meeting, San Francisco, CA, December 2013.